JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 108, NO. D18, 4583, doi:10.1029/2002JD002991, 2003

Optimal parameter and uncertainty estimation of a land surface model: A case study using data from Cabauw, Netherlands

Charles Jackson, Youlong Xia, Mrinal K. Sen, and Paul L. Stoffa

Institute for Geophysics, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA

Received 27 September 2002; revised 29 May 2003; accepted 9 June 2003; published 24 September 2003.

[1] Land surface models involve a large number of interdependent parameters that affect the physics of how surface energy fluxes are partitioned between latent heat, sensible heat, net radiative, and ground heat fluxes. The goal of an optimal parameter and uncertainty analysis of a land surface model is to identify a range of parameter sets that enable model predictions to be bounded within observational uncertainties. Here we apply Bayesian stochastic inversion (BSI) using very fast simulated annealing (VFSA) to identify parameter sets of the Chameleon surface model (CHASM) land surface model that are consistent with the uncertainty limits ascribed to a high-quality data set collected from Cabauw, Netherlands. These results are compared to the parameter sets obtained through the multicriteria (MC) approach. All analyses evaluate model performance against daily and monthly mean observations of sensible, latent, and ground heat fluxes. BSI and MC identify similar "best fit" model parameter sets that improve CHASM performance over default parameter settings. The three most important CHASM parameters at Cabauw are minimum stomatal resistance, vegetation roughness length, and vegetation fraction cover. BSI is based on a Bayesian inference model such that that it expresses uncertainty in terms of a posterior probability density function, different moments of which provide information about parameter means and covariances. Although MC gives a range of possible optimal parameters through the concept of a Pareto set, we found that these ranges did not provide a consistent or representative view of the uncertainty within the observational data. The BSI algorithm in the current study is particularly efficient in that it only requires about double the number of model evaluations than the MC algorithm. This is a substantial saving over other more accurate methods to evaluate uncertainty such as the Metropolis/Gibbs' sampler that requires at least 40 times more computations than the BSI algorithm to obtain similar results. *INDEX TERMS:* 1694 Global Change: Instruments and techniques; 3322 Meteorology and Atmospheric Dynamics: Land/atmosphere interactions; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 9335 Information Related to Geographic Region: Europe; *KEYWORDS:* optimatization, uncertainty, land surface

Citation: Jackson, C., Y. Xia, M. K. Sen, and P. L. Stoffa, Optimal parameter and uncertainty estimation of a land surface model: A case study using data from Cabauw, Netherlands, *J. Geophys. Res.*, *108*(D18), 4583, doi:10.1029/2002JD002991, 2003.

1. Introduction

[2] Land surface modeling is considered one of the major sources of uncertainty for current climate change prediction [*Crossley et al.*, 2000; *Intergovernmental Panel on Climate Change*, 1996]. The Project for Intercomparison of Land surface Parameterization Schemes (PILPS) demonstrates that there is a wide disparity between land surface schemes, with little agreement on predictions of soil moisture and evaporation [*Henderson-Sellers et al.*, 1996]. This disparity comes from differences in model structures (e.g., number of soil layers, number of physical processes represented) and model parameter settings. While the PILPS project gives a qualitative indication of the spread of land surface model predictions created by these differences, it is not clear how representative this spread is of observational uncertainty. Moreover, it is not possible to evaluate how these qualitative uncertainty estimates could be used to estimate uncertainties of climate model predictions that stem from a given land surface model without associating the uncertainty to particular land surface model parameters or knowing how uncertainty in the choice of one parameter affects the choice of other parameters.

[3] We present here Bayesian stochastic inversion (BSI) using very fast simulated annealing (VFSA) [*Sen and Stoffa*, 1995, 1996; C. Jackson et al., An efficient stochastic Bayesian approach to optimal parameter and uncertainty

0148-0227/03/2002JD002991$09.00

estimation for climate model predictions, submitted to *Journal of Climate*, 2002, hereinafter referred to as Jackson, submitted manuscript, 2002] as an efficient, systematic, and quantitative means for identifying optimal parameters and uncertainty for 12 parameters within the Chameleon surface model (CHASM) using data from Cabauw, Netherlands. BSI is an example of an approximate but efficient Monte Carlo importance sampling technique and therefore has as one of its primary objectives the mapping of a multidimensional surface that describes the misfit between observations and model predictions. The application of statistical measures of uncertainty arising from multiple, nonlinearly related parameters has been previously developed for surface hydrology [*Beven and Binley*, 1992; *Beven*, 1993; *Freer et al.*, 1996; *Romanowicz et al.*, 1994, 1996; *Kuczera and Parent*, 1998; *Campbell et al.*, 1999; *Thiemann et al.*, 2001; *Bates and Campbell*, 2001]. The Bayesian formulation of a land surface model problem is most similar to the work of *Franks and Beven* [1997] who provides an example of how Bayesian statistics and Monte Carlo sampling of model parameters can be combined to give likelihood estimates of parameters governing a simple soil vegetation-atmosphere transfer (SVAT) scheme. Franks and Beven find that for a given model performance criterion increasing levels of model complexity leads to a growing number of acceptable model configurations and therefore increasing levels of predictive uncertainty. We expand on this body of work by providing an overview of how parameter optimization is related to uncertainty estimation and the numerical issues related to the efficiency of calculating probability density functions (PPDs) and parameter covariances.

[4] The minima of a multidimensional surface describing the misfit between model predictions and observations as a function of model parameters identify the optimal parameter settings. Identifying these minima is particularly difficult for land surface and hydrology models as many of these models contain more than 50 interdependent parameters. Many methods exist for identifying optimal parameters. Population-based evolution strategies such as the shuffled complex evolution global optimization algorithm have proved to be effective in problems involving simplified rainfall-runoff models [*Wang*, 1991; *Duan et al.*, 1992, 1994]. For more complex models, however, measurement errors and model approximations made it difficult for this method to identify a unique optimal parameter set and led to the development of the multicriteria (MC) method based on the Genetic Algorithm [*Gupta et al.*, 1998]. Recently, *Gupta et al.* [1999] used the MC method to estimate the optimal parameters and uncertainty ranges for a particularly complex land surface model, Biosphere Atmosphere Transfer Scheme (BATS). The results were encouraging in that the BATS model performed substantially better when optimized using the MC method. *Xia et al.* [2002] applied the MC method to the Cabauw data set for the CHASM [*Desborough*, 1999] land surface model to investigate the relationship between the level of model complexity and the accuracy of simulated results. Their results show that complex models perform better than the more simplified models when optimal parameters have been identified. *Sen et al.* [2001] confirmed that the SVAT land surface model that was optimized within "off-line" calculations was able to improve the performance of the Community Climate Model version 3.

[5] The task of sampling the most significant parts of multidimensional misfit between model predictions and observations is distinct from the task of identifying optimal parameters. The latter is much less computationally intense as one need not worry about how one finds the minima of this multidimensional surface. However, it is the shape of the surface around the minima that is most directly related to parameter uncertainty (assuming that the errors are Gaussian). In the case when one may not assume Gaussian errors, PPDs give the best indication of relative likelihood of possible solutions. The BSI methodology strikes a balance between these two objectives. So while we may expect the BSI algorithm to identify a similar set of optimal parameters compared to what may be obtained through the MC methodology, we may also expect that BSI will involve more model evaluations to obtain the additional information needed to describe the uncertainty. Our main objective is to apply the BSI to the identification of optimal parameters and uncertainty estimates for the CHASM land surface model using the Cabauw data set. Besides the work by *Franks and Beven* [1997] who discuss the Bayesian formulation of uncertainty estimation in a general sense, we are not aware of any previous attempt to evaluate and display uncertainties for land surface model parameters. We also compare optimal parameters and uncertainty ranges obtained through the BSI analysis with the optimal parameters and estimates of the range of optimal parameters obtained through the MC methodology. The purpose of this comparison is to provide a consistency check on our results as well as to examine more closely the meaning of optimal parameter uncertainty as may be seen through the Pareto set (defined to be the set of acceptable solutions when no weighting between multiple criteria is specified) that is produced by the MC method.

2. Cabauw Data and CHASM Land Surface Model

[6] The observations of land surface energy fluxes at Cabauw (51°58′N, 4°56′E), Netherlands, are described in detail by *Beljaars and Bosveld* [1997]. The quality of the data is considered to be very good. Measurements were made from a tower surrounded by short grass divided by narrow ditches without obstacles or perturbation of any importance within a distance of 200 m. Beyond 200 m some scattered houses and trees can be found. The climate in the area is characterized as moderate maritime with prevailing westerly wind. Data were collected at half hour intervals for the entire year of 1987. During this year, 776 mm of precipitation fell (including several snowfall events) and the annual mean air temperature was 282 K. The vegetation cover is close to 100% year round. The soil contains 35–55% clay. Measurements were taken of downward short-wave radiation, downward long-wave radiation, precipitation, air temperature, wind speed, specific humidity, sensible heat flux (SH), latent heat flux (LH), net radiation (R_{net}), and ground surface heat flux (GS). The Cabauw data set has been used to investigate relative performance differences of several commonly used land surface models [*Chen et al.*, 1997; *Desborough*, 1999; *Xia et al.*, 2002].

[7] CHASM is designed to accommodate a wide range of representations of the land surface energy balance within a

common modeling framework [*Desborough*, 1999]. The representation of the land surface can take on a number of formulations ranging from a complex mosaic structure [*Koster and Suarez*, 1992] all the way to the most simple zero energy balance formulation [*Manabe*, 1969]. Here we use only the complex mosaic representation. Within this representation, the land-atmosphere interface is divided into two tiles. The first tile is a combination of bare ground and exposed snow, with the second tile consisting of dense vegetation. The tiles may be of different sizes, and the fluxes out of each tile are area-weighted. Because a separate surface energy balance is calculated for each tile, temperature variations may exist across the land-atmosphere interface. Depending on the surface type, a prognostic bulk temperature for the storage of energy and a diagnostic skin temperature for the calculation of surface energy fluxes are determined. Snow cover fraction for both the ground and foliage surface is calculated as functions of the snowpack depth, snowpack density, and vegetation roughness length. The vegetation fraction is divided into wet and dry fractions if canopy interception is allowed. This model has explicit parameterizations for canopy resistance, canopy interception, vegetation transpiration, and bare ground evaporation but has no explicit canopy air space.

[8] CHASM uses the formulation of *Manabe* [1969] for the hydrologic component of the land surface in which the root zone is treated as a bucket with finite water holding capacity. Any water accumulation beyond this capacity is assumed to be runoff. Except for moisture in the root zone, water can be stored as snow on the ground or canopy. The soil contains four layers with finite heat capacity. The calculation of soil temperature assumes no heat flux across the bottom of the lowest layer. Each tile has four evaporation sources: canopy evaporation, transpiration, bare ground evaporation, and snow sublimation. More details on CHASM can be found in the work of *Desborough* [1999] and *Xia et al.* [2002].

3. Methods of Optimal Parameter and Uncertainty Estimation

3.1. Bayesian Stochastic Inversion

[9] The goal of BSI is to estimate a multidimensional (joint) probability distribution that expresses which combination of model parameter values are deemed "acceptable." It is not logical or possible to express this probability in an absolute sense. Rather, it is necessary to express these probabilities as being conditional on certain desired constraints. Specifically, we would like to know which sets of model parameter values would allow model predictions to exist within observational uncertainties. The Bayesian formulation expresses the basic concept of conditional probabilities that we are interested in, where the posterior PPD is the desired result. According to Bayes' rule, derived from the definition of conditional probabilities, the PPD is defined as

$$\sigma(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \frac{l(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})}{\int l(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m})d\mathbf{m}}, \tag{1}$$

where $\mathbf{m}$ is a vector containing a set of parameter values $\mathbf{d}_{\text{obs}}$ is the data vector, $\sigma(\mathbf{m}|\mathbf{d}_{\text{obs}})$ is the conditional probability for each set of parameter values represented by vector $\mathbf{m}$ (given data in vector $\mathbf{d}_{\text{obs}}$), $l(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is the likelihood function expressing the conditional probability for reproducing observations $\mathbf{d}_{\text{obs}}$ from a model with parameter values $\mathbf{m}$, $p(\mathbf{m})$ is a "prior" probability for $\mathbf{m}$ (given expert judgment or other reasons to constrain the possible choices of $\mathbf{m}$ independent of data $\mathbf{d}_{\text{obs}}$). Because the data vector $\mathbf{d}_{\text{obs}}$ within the likelihood function does not change, it should be understood that the likelihood function is primarily a function of $\mathbf{m}$ [*Box and Tiao*, 1992]. When Gaussian errors are assumed within the observations and model predictions, the likelihood function takes the form,

$$l(\mathbf{d}_{\text{obs}}|\mathbf{m}) \propto \exp(-S \cdot E(\mathbf{m})), \tag{2}$$

where $E(\mathbf{m})$ is a cost function that gives some measure of mismatch between observations and model predictions and S is a scaling factor discussed more fully in section 4.3. The cost function can be defined in many ways. Assuming Gaussian errors in the data, the most appropriate choice is

$$E(\mathbf{m}) = \sum_{i=1}^{N} \frac{1}{2N} \left[(\mathbf{d}_{\text{obs}} - g(\mathbf{m}))^T \mathbf{C}^{-1} (\mathbf{d}_{\text{obs}} - g(\mathbf{m})) \right]_i, \tag{3}$$

where there are N different sets of observations, $g(\mathbf{m})$ is the forward model, and $\mathbf{C}^{-1}$ is the inverse of the data covariance matrix that includes both the observational error and modeling uncertainty error. Although the form of the PPD based on equation (3) assumes Gaussian errors in observations and in model predictions, because the forward model is a part of the error function, there is no expectation that the PPD would itself be Gaussian [*Tarantola*, 1987].

[10] Once the PPD is known, the parameter means $\langle \mathbf{m} \rangle$ or covariances can be obtained through multidimensional integrals of the general form

$$\mathbf{I} = \int f(\mathbf{m}) \sigma(\mathbf{m}|\mathbf{d}_{\text{obs}}) d\mathbf{m}, \tag{4}$$

where $f(\mathbf{m}) = \mathbf{m}$ for the parameter means or $f(\mathbf{m}) = (\mathbf{m} - \langle \mathbf{m} \rangle)(\mathbf{m} - \langle \mathbf{m} \rangle)^T$ for the parameter covariance matrix. Because the PPD is multidimensional, it is difficult to visualize. One approach is to display the marginal PPD, defined to be the one-dimensional projection of the multidimensional PPD (equivalent to $f(\mathbf{m}) = 1$ in equation (4) and integrating over all dimensions except for the dimension of interest).

3.2. Numerical Methods

[11] The principle challenge in calculating integrals based on equation (4) is in deriving the PPD, which can be very intense computationally. We review here several numerical methods for estimating the PPD and focus on VFSA which has been highlighted by *Sen and Stoffa* [1996] as being particularly efficient. Jackson et al. (submitted manuscript, 2002) provide a more complete description of VFSA and how it may be optimized for particular problems.

3.2.1. Grid Search

[12] This straightforward method involves subdividing model parameter space into a number of equally spaced intervals and enumerating every possible combination of

model parameters and evaluating the cost function for each of these combinations [*Sen and Stoffa*, 1996]. The disadvantages of this method are the large number of forward model calculations, many of which do not contribute substantially to the integral in equation (4), and the fact that the resolution is constrained by the interval spacing.

3.2.2. Gibbs' Sampler

[13] The Gibbs' sampler is a version of an "importance sampling" technique that improves the efficiency of the calculation by sampling model parameter sets from the Gibbs' distribution which is, in effect, equivalent to the desired PPD [*Metropolis et al.*, 1953; *Kirkpatrick et al.*, 1983]. The concept of the Gibbs' sampler originates from the numerical techniques developed in statistical mechanics to simulate the macroscopic behavior of a system with a large number of interacting particles. The numerical implementation of the Gibbs' sampler is based primarily on the heat-bath algorithm in which the relative probability of different model parameter sets are evaluated in advance of any model evaluations [*Geman and Geman*, 1984; *Rothman*, 1986]. This algorithm also requires parameter space to be subdivided into a number of equally spaced intervals. Another variant of the Gibbs' sampler is based on the Metropolis algorithm [*Metropolis et al.*, 1953; *Kirkpatrick et al.*, 1983]. In the Metropolis formulation, a starting model is selected at random and the cost function is evaluated $E(\mathbf{m}_i)$. The starting model is perturbed to obtain a new model $\mathbf{m}_{i+1}$ and new cost function evaluation $E(\mathbf{m}_{i+1})$. If the change in "energy" $\Delta E = E(\mathbf{m}_{i+1}) - E(\mathbf{m}_i)$ is negative, then the new model is accepted. If the change is positive, the new model is accepted with a probability

$$P = \exp\left(\frac{-\Delta E}{T}\right), \tag{5}$$

where T is a control parameter analogous to temperature. After a large number of iterations at constant temperature, the pdf for $\mathbf{m}$ (notated as *prob*($\mathbf{m}$)) converges to the following Gibbs' distribution

$$\mathrm{prob}(\mathbf{m}) = \frac{\exp(-E(\mathbf{m})/T)}{\sum \exp(-E(\mathbf{m})/T)}. \tag{6}$$

[14] As noted by *Sen and Stoffa* [1996], the fact that the distribution for prob($\mathbf{m}$) with $T = 1$ in equation (6) is the same as for $\sigma(\mathbf{m}/\mathbf{d}_{\mathrm{obs}})$ in equation (1), except for the presence of the prior distribution $p(\mathbf{m})$, means that when using the Gibbs' sampler at constant temperature, the distribution will converge to the distribution of the PPD without bias when assuming a constant prior. Because sampled models are not discretized in the Metropolis formulation of the Gibbs' sampler, the resolution of the Metropolis/Gibbs' sampler will depend on the number of iterations. The resolution in general will be much greater near the minimum in $E(\mathbf{m})$ because these regions tend to be sampled more often. While this method can substantially improve the efficiency of evaluating the PPD, the number of forward model calculations is still prohibitively large for many purposes.

3.2.3. Very Fast Simulated Annealing

[15] One may use the temperature construct within the Metropolis algorithm to locate the global minimum in $E(\mathbf{m})$ by very slowly lowering the temperature parameter within equation (6). This process is analogous to the annealing process within a physical system, whereby the lowest energy state between atoms or molecules (the crystalline form) is achieved by gradually cooling the substance within a heat bath. Because of this physical analogy, the algorithm is known as Simulated Annealing. *Ingber* [1989] introduced within the Simulated Annealing algorithm a new procedure for selecting parameter sets according to a temperature-dependent Cauchy distribution. This modified algorithm further enhances the ability of Simulated Annealing to converge efficiently and robustly to a minimum in $E(\mathbf{m})$ and is referred to as VFSA. The acceptance criterion for successive model selections is the same as for the Metropolis rule.

[16] One of the central points in the discussion on what efficiencies can be afforded within the numerical methods that are adopted for optimal parameter and uncertainty estimation is that there needs to be a balance between the computational effort that is spent identifying optimal model parameters and the computational effort that is spent mapping the multidimensional PPD. The VFSA algorithm as presented by *Ingber* [1989] and used by *Sen and Stoffa* [1996] is an efficient method to identify optimal parameters, especially when nonlinearities are important. At the other extreme, Monte Carlo or Grid search algorithms would provide the most accurate (nonbiased) map of the multidimensional PPD but may require more model evaluations than one can afford to make. *Sen and Stoffa* [1996] argue that one can slightly adapt the VFSA cooling schedule, convergence acceptance criterion, and allow for numerous repetitions of the minimization procedure to strike a balance between these two objectives that is both efficient and effective. Depending on the application, one can save several orders of magnitude in the number of model evaluations over either the Monte Carlo or Grid search algorithms [*Sen and Stoffa*, 1996].

[17] The PPD derived through the VFSA search algorithm is unavoidably biased toward the peaks of the PPD through its procedure to change the temperature control parameter during the selection process of model parameters. However, the VFSA algorithm may be repeated a number of times with different starting models to allow sufficient sampling of the entire model space. This reduces the biases of the PPD and improves the estimates of the model covariance matrix. While variances may be underestimated relative to what may be obtained through the Metropolis/Gibbs' sampler, the normalized covariance matrix (the correlation matrix) has been found to be nearly equivalent between the two approaches [*Sen and Stoffa*, 1996].

3.3. Multicriteria Approach

[18] The MC parameter estimation methodology is primarily developed as an optimization procedure that uses a Genetic Algorithm to search out parameter sets that improve a model's performance [*Yapo et al.*, 1998; *Gupta et al.*, 1998]. The term "multicriteria" refers to the fact that the cost function is a vector giving a single cost function measure for each field used to evaluate a model's performance. The cost function displayed in equation (3) is considered a single criterion as the mismatch between

Table 1. The Names, Default Values, Optimal Values Determined Through the Multicriteria Approach and Very Fast Simulated Annealing, Ranges, and Descriptions of 13 Parameters of the CHASM Land Surface Model[a]

Parameter	Default	MC	VFSA	Minimum	Maximum	Description
ALBG	0.20	0.24	0.39	0.10	0.40	bare ground albedo
ALBN	0.75	0.76	0.98	0.50	1.00	snow albedo
ALBV	0.23	0.27	0.26	0.10	0.40	vegetation albedo
LEFM	2.00	5.0	4.5	4.0	6.0	maximum leaf area index
LEFS (fixed)	1.70	1.7	1.7	1.7	1.7	leaf area index seasonality parameter
VEGM	0.95	0.96	0.99	0.70	1.00	maximum fractional vegetation cover
VEGS	0.25	0.25	0.23	0.23	0.26	fractional vegetation cover seasonality
RCMIN	40.0	17.8	16.7	1.0	300	minimum canopy resistance, s/m
WRMAX	141	292	130	10	600	available water holding capacity, mm
Z0G	1.00	0.2	0.94	0.01	1.00	bare ground roughness length (10^{-2} m)
Z0N	4.00	4.00	1.73	1.00	6.00	snow surface roughness length (10^{-4} m)
Z0V	0.15	0.02	0.02	0.01	2.5	vegetation roughness length, m
T_S	279	286	282	275	310	initial surface temperature, K

[a]MC, Multicriteria approach; VFSA, very fast simulated annealing.

model and observations for multiple fields is ultimately expressed as a single number where one must provide a reasonable weighting between each of the N fields used to evaluate model performance. In the case of equation (3) model-data comparisons of fields with different units are weighted using the covariance matrix. The MC method tries to avoid making assumptions about what the appropriate weighting should be and instead identifies the Pareto set, which is the set of model parameters that minimize the mismatch between model predictions and observations for each of the separate criteria. However, when it comes time to picking an optimal parameter set from within the Pareto set, one must choose an appropriate weighting anyway. The optimal parameter sets that we select from the Pareto set assumes equal weighting among all criteria.

4. Cost Function

4.1. Definition of Cost Function

[19] CHASM contains 13 parameters. Although many parameters are not critical to the land surface energy balance at Cabauw, we consider all parameters within the BSI and MC analyses. The range of possible values and default settings for each of these parameters are shown in Table 1. The Cabauw data set contains four fields that can be compared with model predictions including SH, LH, ground heat flux (GS), and R_{net}. Because the information about R_{net} is included in the forcing of the model, R_{net} is not independent of model evaluations of SH and LH, and therefore we choose not to use R_{net} to evaluate model performance. The measure of mismatch between observations and model predictions is given by equation (3). For BSI, we keep only the diagonal components of the covariance matrix (variances) and use the Cabauw data set to evaluate the variance for each half-hour segment of the day (i.e., covariance matrix collapses into a vector with 48 values). For the MC calculation, we reduce the dimensionality of the covariance matrix further and use only average variance for all data in the Cabauw data (i.e., covariance matrix collapses into a scalar). This latter definition of the cost function has been used in previous applications of the MC method [*Xia et al.*, 2002].

4.2. Cost Function Profiles

[20] Before attempting any optimal parameter or uncertainty estimation of the CHASM model, it may be helpful to use cost function "profiles" to display which parameters are likely to be important. The cost function profiles can also be used to objectively select parameters that most affect the uncertainty. Conducting an uncertainty analysis over a reduced number of parameters can significantly reduce the cost of deriving the multidimensional PPDs. A similar procedure is advocated by *Bastidas et al.* [1999] for reducing the number of parameters involved in the MC optimal parameter analysis.

[21] A cost function profile is a graph of the cost function as a function of variations in a given parameter while holding the value of all other parameters constant. These graphs provide an expectation of the relative sensitivity of the model to each of the model parameters. Because this evaluation assumes linear independence between model parameters, this estimate of the sensitivity may not reflect the model's true sensitivity. However, if the model's response to variations in model parameters were truly linear, then the cost function profiles would provide an accurate representation of the multidimensional PPD with only a few model evaluations.

[22] Figure 1 shows a cost function profile analysis for 12 of the CHASM parameters. As we do not know in advance

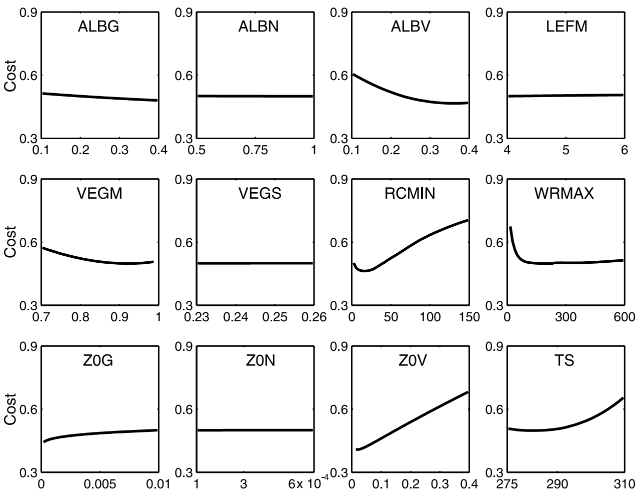

Figure 1. Cost function profiles for 12 parameters within the CHASM land surface model.

what parameter values would be optimal, we use the default values for the parameters that are being held fixed. The default values are the best guess for the parameter values based on expert opinion and what can be inferred from observations. The results show that the most important parameters are minimum stomatal resistance (RCMIN) and vegetation roughness length (Z0V). In addition, vegetation albedo (ALBV), vegetation fraction (VEGM), maximum soil moisture holding capacity (WRMAX), and initial soil temperature (T_S) have some influence on simulation errors. The other parameters such as ground albedo (ALBG), snow albedo (ALBN), maximum leaf average index (LEFM), vegetation fraction cover seasonality (VEGS), ground roughness length (Z0G), and snow roughness index (Z0N) have negligible effect on simulation errors. Because Cabauw is nearly completely covered by vegetation for all seasons, the parameters that end up being most important are related to vegetation. Other parameters such as ALBG and Z0G, which are related to bare soil, and VEGS, which is concerned with the seasonal variation of vegetation fraction cover, have little effect on simulation errors.

[23] The parameters that are important are partly a function of the observations that are used to evaluate model performance. For instance, snow albedo (ALBN) may have affected the cost function more if observations of snow depth and snowmelt were available and used within the definition of the cost function. In contrast, minimum stomatal resistance (RCMIN), vegetation roughness length (Z0V), and the vegetation fraction (VEGM) can be more directly related to sensible and LHs.

4.3. Specification of Cost Function Scaling Factor

[24] The multidimensional PPD is intended to provide information about the relative likelihood of different choices of model parameters. Obviously, if there is a lot of uncertainty within the observations or model predictions, it would be very difficult to state confidently that one parameter set is significantly better than any other. In this case the PPD should look rather flat, giving nearly equal weight to all parameters. While the covariance matrix within the cost function could give an appropriate weighting and thus scaling of the relative likelihood between alternate parameter choices, often we do not have complete knowledge of this uncertainty. In the case of the Cabauw data, the published uncertainty estimates [*Beljaars and Bosveld*, 1997] are given by monthly means rather than in terms of errors in the half-hourly measurements that are used within the cost function. We can use these published monthly mean error estimates to help determine an appropriate scaling factor that will shape the probability distributions. After one has mapped the multidimensional cost function surface using BSI, one may rank all the models according to their cost. The lowest value of the cost function ($E_{\min}$) corresponds to the cost of the optimal parameter set. Given estimates of the errors in the data we can identify a cost function threshold (E_{thresold}) that is representative of the uncertainty in the data. In the case of the Cabauw data set, we assume that the published error estimates of ±5 and ±10 W/m^2 for monthly mean sensible and LHs are estimates of the two-sigma error (the 95% confidence interval). We identified the cost function threshold ($E_{\mathrm{threshold}}$) that specifies when 95% of the selected model parameter sets are within ±5 W/m^2 of the SHs of the optimal model prediction. One could have gone through this exercise using published uncertainty estimates of the LHs; however, we found that we obtained nearly the same answer. We then apply the logic that parameter sets that are ΔE (ΔE = $E_{\mathrm{threshold}} - E_{\min}$) away from the optimal parameter set will be given a likelihood measure (equation (2)) of exp(-2), which is equivalent to the 95% probability measure for a Gaussian distribution. This implies $S = 2/\Delta E$.

[25] For the BSI analysis of the Cabauw data and particular definition of the cost function, we found that 46.7 is the most appropriate value for the scaling factor for the BSI analysis based on VFSA and 76.7 is the appropriate value for the analysis using the Gibbs' sampler. One may confirm that this provides an adequate description of the uncertainty later on when we show that the 95% confidence interval encompasses most of the observed monthly mean sensible and LHs (see section 5). The appropriateness of the scaling factor for the Gibbs' sampler is less certain than the one for VFSA since there were so few parameter sets included within the uncertainty of observations (only 57 sets for the Gibbs' sampler as opposed to 34,080 sets for VFSA).

5. Results

[26] The marginal PPD gives some indication of the shape of the multidimensional PPD for particular parameter "dimensions" of interest. Shown in Figure 2 is the approximate marginal PPD derived from the BSI methodology based on VFSA (solid) and the Gibbs' sampler (dashed). The number of forward model integrations required to generate stable distributions differed between these two approaches, with VFSA requiring substantially fewer iterations (24,000) than the Gibbs' sampler (942,000). Moreover, the BSI analysis based on VFSA was able to identify over 600 times the number of parameter sets within observational uncertainties relative to the Gibbs' sampler.

[27] The multidimensional PPD has been normalized by its volume integral. From this scaling one may infer the relative importance of each parameter by the vertical scaling of the marginal PPDs. For parameters that are the most important (RCMIN, Z0V, VEGM) the probability distribution estimated by VFSA is quite similar to that estimated by the Gibbs' sampler. The differences that are apparent between these two distributions are more evident within parameters that may be considered less important. Less weight should be given to specific features within these distributions as individual peaks may reflect artifacts of the sampling rather than real features of the cost function.

[28] The circles in Figure 2 show the optimal parameters that were identified using BSI. The circles often line up with the peaks in the marginal PPD, although this is not always the case since there is no requirement that optimal parameters are also the most probable. The probability assigned to a particular parameter value through the PPD involves a combined measure of the likelihood function and the frequency at which parameter values within a given neighborhood are selected. There also can be differences that arise from the fact that we are looking at a one-

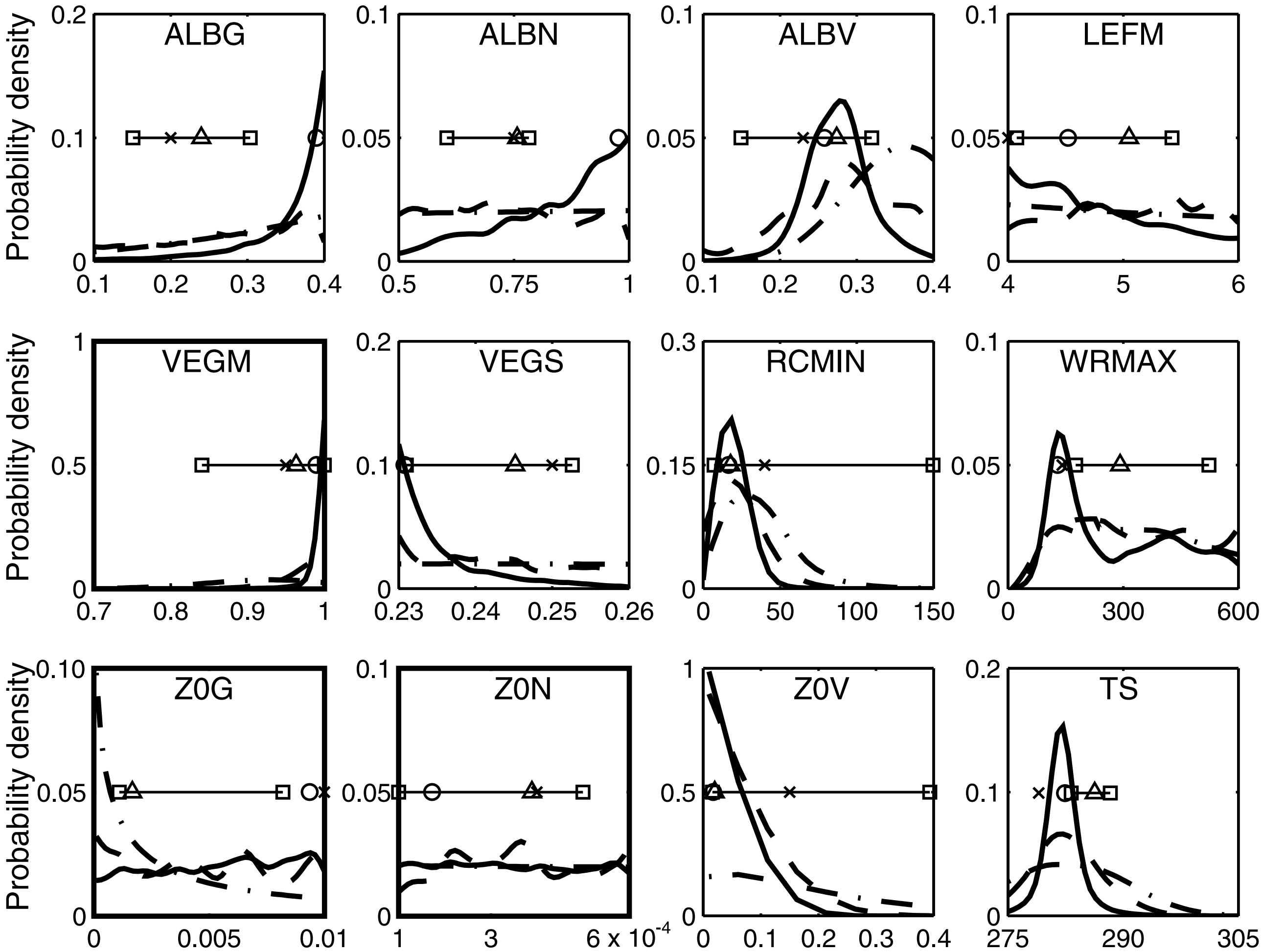

Figure 2. Marginal PPD results from analyses based on VFSA (solid line), Gibbs' sampler (dashed line), and cost function profiles (dash-dotted line). The two boxes joined by a line indicate the Pareto set from the MC analysis. Also shown are the best performing parameter sets for VFSA (circles) and MC (triangles). The default parameter set is indicated by a cross.

dimensional projection of a multidimensional PPD. The optimal parameters that were identified through the MC analysis are indicated by triangles. For many of the parameters, the optimal set obtained through the BSI and MC methods are similar, especially for the parameters that are the most influential. These optimal parameter sets can differ substantially from the default parameter set (indicated by a cross).

[29] Also shown in Figure 2 is the PPD one would obtain from the linear cost function profiles of Figure 1. The difference between the linear calculation and the probability distributions based on either of the VFSA or the Gibbs' sampler gives one indication of the effects of model interdependences on joint probability distributions.

[30] The parameters that were expected to be important through consideration of the cost function profiles (RCMIN, Z0V) are among the parameters that have been identified through the BSI analysis to be influential. However, the BSI analysis also reveals that VEGM increased its relative influence as compared to what was expected from the cost function profile analysis. One potential explanation for this is that VEGM has a correlation of −0.26 with Z0V (Table 2), one of the two most important parameters, and that this association increased VEGM's relative influence (note that for parameters that are linearly dependent, a significant correlation is usually reserved for correlation values of 0.4 or higher. However, when nonlinearities may be important the true significance of interparameter dependencies may be underrepresented by the correlation coefficient.) Moreover, the cost function profiles were created using the default parameter settings. The default setting for Z0V is significantly larger than what the BSI analysis shows to be the most favorable setting. This "error" may have affected our perception of the true importance of percent vegetation cover (VEGM). Something similar may be affecting the relative importance of ALBV and T_S. The cost function profiles would suggest that ALBV should be more important than T_S as the cost function reaches a lower minimum for ALBV than T_S. The correlation matrix shows that T_S is not strongly correlated with any other parameter, which suggests that something must be reducing the relative importance of ALBV. In this case vegetation albedo (ALBV) is strongly correlated with minimum canopy resistance (RCMIN) and the slight overestimate in RCMIN's default setting may have unduly given too much weight to ALBV.

Table 2. Correlation Matrix of 12 of the CHASM Land Surface Model Parameters[a]

Parameter	ALBG	ALBN	ALBV	LEFM	VEGM	VEGS	RCMIN	WRMAX	Z0G	Z0N	Z0V	T_S
ALBG	1.00											
ALBN	0.09	1.00										
ALBV	−0.16	−0.05	1.00									
LEFM	−0.08	−0.08	0.06	1.00								
VEGM	0.11	0.14	−0.07	−0.06	1.00							
VEGS	−0.10	−0.14	0.00	0.08	−0.13	1.00						
RCMIN	−0.02	−0.07	**0.54**	0.09	−0.11	0.06	1.00					
WRMAX	−0.09	−0.06	−0.11	0.02	−0.03	0.05	−0.14	1.00				
Z0G	0.07	0.06	−0.05	−0.02	0.11	0.00	−0.06	0.01	1.00			
Z0N	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00		
Z0V	−0.14	−0.13	0.05	0.01	−**0.26**	0.12	0.09	0.04	−**0.33**	0.00	1.00	
T_S	−0.03	0.00	−0.03	−0.03	−0.01	0.03	−0.01	0.01	−0.01	0.00	0.04	1.00

[a]Correlations >0.25 are bolded. ALBG, ground albedo; ALBN, snow albedo; ALBV, vegetation albedo; LEFM, maximum leaf average index; VEGM, vegetation fraction; VEGS, vegetation fraction cover seasonality; RCMIN, minimum stomatal resistance; WRMAX, maximum soil moisture holding capacity; Z0G, ground roughness length; Z0N, snow roughness index; Z0V, vegetation roughness length; T_S, initial soil temperature.

[31] The uncertainty in the specification of the optimal parameters between the BSI and MC methods is quite different. The range of acceptable solutions indicated by the Pareto set (indicated in Figure 2 by a line and two boxes) is nearly unbounded for the two most important parameters RCMIN and Z0V. If, however, we left observations of ground heat flux out of the definition of the cost function, the range of the Pareto set reduces to nearly a

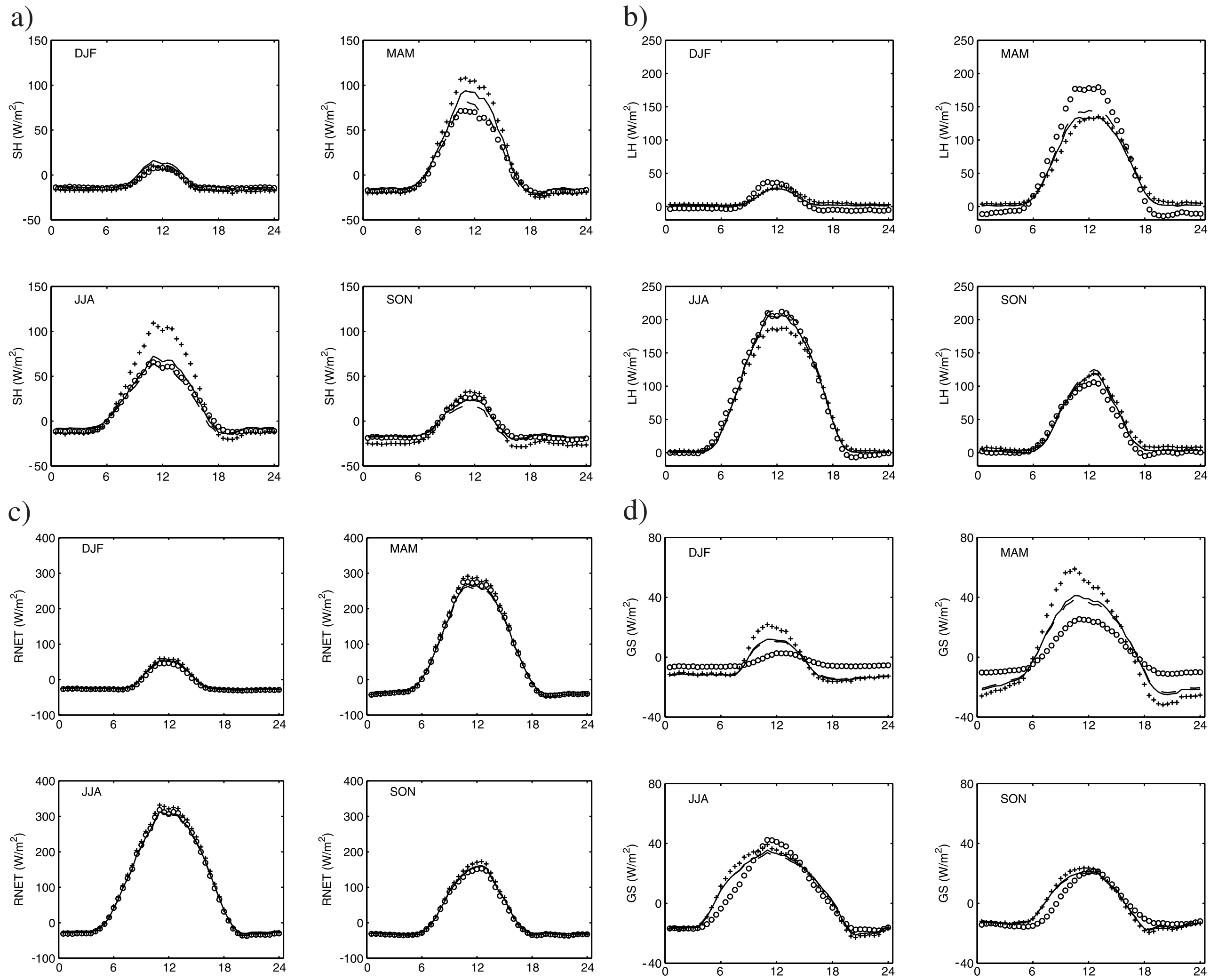

Figure 3. Seasonally averaged diurnal cycle of (a) sensible heat, (b) LHs, (c) net radiative fluxes, and (d) ground heat fluxes for observations (circles), default parameter set (plus), best parameter set using VFSA (solid line), and best parameter set using MC (dashed line).

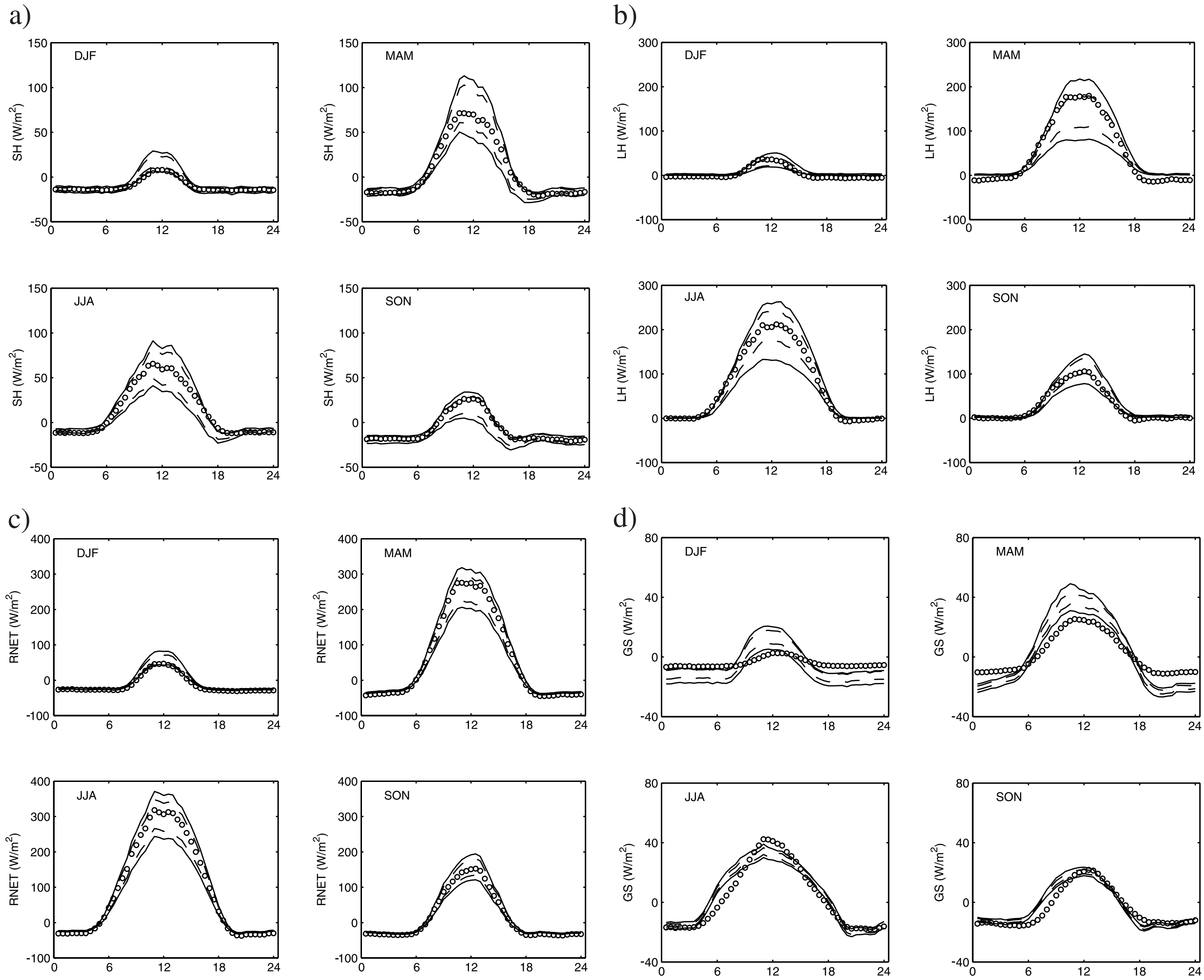

Figure 4. Seasonally averaged diurnal cycle of (a) SHs, (b) LHs, (c) net radiative fluxes, and (d) ground heat fluxes for observations (circles) and outer envelope of CHASM land surface model predictions using parameter sets included within the 95% (solid line) and 60% (dashed line) probability volume of the multidimensional PPD.

single point (not shown). This is because minimum canopy resistance (RCMIN) and vegetation roughness length (Z0V) have little impact on ground heat fluxes and therefore are not well constrained by these observations. As the Pareto set is defined from what would minimize the cost function when one is uncertain about the relative weighting between multiple criteria, it becomes clear that the Pareto set shows what the maximum uncertainty is for the observations that least constrain any given parameter. Even if one were to leave observations of ground heat flux out of the definition of the cost function, the fact that the acceptable solutions are reduced to nearly a single point indicates that in this case the Pareto set could also underestimate the uncertainty. One must keep in mind that the main purpose of the MC method is to estimate optimal parameter sets. The optimal parameters obtained using the MC method are nearly identical to those obtained from BSI. The average monthly mean SHs for both methods were found to be within 2.5 W/m^2 of observations. However, with BSI using VFSA, we were able to obtain accurate estimates of the parameter uncertainties with only twice the number of forward model evaluations as compared to the MC methodology.

[32] Another way to visualize the performance of the optimal parameters and their uncertainty is to display the seasonally averaged diurnal cycle comparison between observations and model predictions. The performance of the optimal parameters using BSI (solid), MC (dashed), and the default parameter set (plus) as compared to observations of SHs, LHs, net radiation, and ground heat flux (circles) is shown in Figures 3a–3d. The predictions of optimal parameters are nearly identical for all fields during all seasons and are substantially closer to observations than the default parameter set. There is close agreement between model predictions and observations of SHs and net radiation, fairly good agreement with observations of LHs for all seasons except spring, and particularly poor agreement with ground heat fluxes during winter and spring. A similar mismatch between simulations and observations can be found in the work of *Chen et al.* [1997]. The poor match to ground heat fluxes may in part be due to errors in the observations which must be inferred from measured temperature differences in

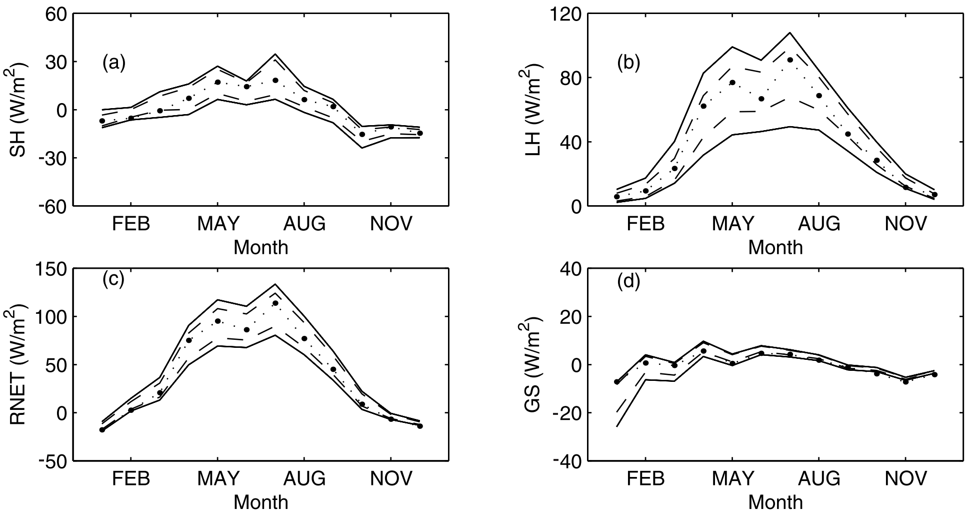

Figure 5. Monthly mean (a) SHs, (b) LHs, (c) net radiative fluxes, and (d) ground heat fluxes for observations (dots) and outer envelope of CHASM land surface model predictions using parameter sets included within the 95% (solid line) and 60% (dashed line) probability volume of the multidimensional PPD.

the soil between 0- and 2-cm depth or extrapolated from heat fluxes at 5- and 10-cm depth. However, there are known inadequacies in the formulation of snow evaporation that may account for some of the large differences that occur during winter and spring [*Slater et al.*, 2001].

[33] The outer envelope for model predictions using parameter sets within the 95 and 60% confidence intervals for the seasonally averaged diurnal cycle for each of the observations are displayed in Figures 4a–4d. Figure 5 shows these same envelopes and a comparison to observations for predictions of monthly means. The number of parameter sets that are included in the 95 and 60% confidence intervals is 34,080 and 13,475, respectively, out of a possible total of 97,311. The number of models within each of these ranges depends on the total number of forward model evaluations used within the analysis. Depending on how well one needs to describe the probability distribution of a given confidence interval, one may need to pick a random sampling from these parameter sets or continue the BSI analysis until a suitable number has been reached. The 95% confidence interval includes observations for all seasons and for all observations except ground heat flux during winter and spring and part of the morning hours during summer and fall. Moreover, the 95% confidence interval includes most of the predictions of the default model except for SHs during summer, and ground heat fluxes during winter and spring. The joint probability of the default parameter set is 4%, which places it just outside the two-sigma uncertainty limit and nominally could be excluded as a realistic combination of model parameters. This evaluation, however, is relative to the optimal parameter set which has its own deficiencies (particularly with its predictions of ground heat flux).

6. Summary and Conclusions

[34] We have presented a new approach to optimal parameter and uncertainty estimation using BSI. We have applied this analysis to the Cabauw data set in order to constrain 12 parameters within the CHASM land surface model. The primary objective of the BSI analysis is to derive a multidimensional PPD, which can be an intense computation. We show that a BSI analysis using VFSA can reduce the number of required model evaluations by over 40 times what is required by the Gibbs' sampler, with little difference in derived marginal probability distributions for individual model parameters. We have compared the ability of BSI and MC methods to identify the optimal parameters and found they give nearly identical results but that estimates of uncertainty differed substantially. Although MC gives a range of possible optimal parameters through the concept of a Pareto set, we found that these ranges did not provide a consistent or representative view of the uncertainty within the observational data.

[35] The three most important parameters of the CHASM model are minimum stomatal resistance, vegetation roughness length, and vegetation fraction cover. The importance of these parameters is likely due to the fact that Cabauw is nearly 100% covered by vegetation year round. We also observed that there are important relationships between model parameters that cause an uncertainty in one parameter to affect the uncertainty in other parameters. Because of this, it is important that when choosing parameter sets to represent the predictive uncertainty of the CHASM model within climate model simulations, the parameter sets need to reflect the interdependencies between model parameters. These sets may either be obtained directly from the sets that make up the multidimensional PPD or estimated from the mean and covariance matrix derived from the PPD if and only if the PPD itself is well approximated by a Gaussian distribution.

[36] **Acknowledgments.** The G. Unger Vetlesen Foundation supported C.J. and Y.X. Special thanks go to Hoshin Gupta for providing the MC methodology and Andy Pitman for providing the CHASM land surface model. We acknowledge the Royal Netherlands Meteorological Institute for providing the Cabauw data set.

References

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Sensitivity analysis of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490, 1999.

Bates, B. C., and E. P. Campbell, A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*, 937–947, 2001.

Beljaars, A. C. M., and F. Bosveld, Cabauw data for the validation of land surface parameterization schemes, *J. Clim.*, *10*, 1172–1193, 1997.

Beven, K. J., Prophecy, reality and uncertainty in distributed hydrological modeling, *Adv. Water Resour.*, *16*, 41–51, 1993.

Beven, K. J., and A. M. Binley, The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*, 279–298, 1992.

Box, G., and G. Tiao, *Bayesian Inference in Statistical Analysis*, 588 pp., John Wiley, Hoboken, N. J., 1992.

Campbell, E. P., D. R. Fox, and B. C. Bates, A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, *Water Resour. Res.*, *35*, 211–220, 1999.

Chen, T. H., et al., Cabauw experimental results from the Project for Intercomparison of Land-surface Parameterization Schemes, *J. Clim.*, *10*, 1194–1215, 1997.

Crossley, J. F., J. Polcher, P. M. Cox, N. Gedney, and S. Planton, Uncertainties linked to land-surface processes in climate change simulations, *Clim. Dyn.*, *16*, 949–961, 2000.

Desborough, C. E., Surface energy balance complexity in GCM land surface models, *Clim. Dyn.*, *15*, 389–403, 1999.

Duan, Q., S. Sorooshian, and V. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*, 1015–1031, 1992.

Duan, Q., S. Sorooshian, and V. K. Gupta, Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*, 265–284, 1994.

Franks, S. W., and K. J. Beven, Bayesian estimation of uncertainty in land surface-atmosphere flux predictions, *J. Geophys. Res.*, *102*(D20), 23,991–23,999, 1997.

Freer, J., K. J. Beven, and B. Ambroise, Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*, 2161–2173, 1996.

Geman, S., and D. Geman, Stochastic relaxation, Gibbs' distribution, and bayesian restoration of images, *IEEE Trans.*, *PAMI-6*, 721–741, 1984.

Gupta, H. V., S. Sorooshian, and P. O. Yapo, Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–761, 1998.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang, Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,503, 1999.

Henderson-Sellers, A., K. McGuffie, and A. Pitman, The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS): 1992–1995, *Clim. Dyn.*, *12*, 849–859, 1996.

Ingber, L., Very fast simulated re-annealing, *Math. Comput. Modell.*, *12*(8), 967–973, 1989.

Intergovernmental Panel on Climate Change, *Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., 572 pp., Cambridge Univ. Press, New York, 1996.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science*, *220*, 671–680, 1983.

Koster, R. D., and M. J. Suarez, Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *J. Geophys. Res.*, *97*(D3), 2697–2715, 1992.

Kuczera, G., and E. Parent, Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*, 69–85, 1998.

Manabe, S., Climate and the ocean circulation, I, The atmospheric circulation and the hydrology of the Earth's surface, *Mon. Weather Rev.*, *97*, 739–805, 1969.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem Phys.*, *21*, 1087–1092, 1953.

Romanowicz, R., K. J. Beven, and J. A. Tawn, Evaluation of predictive uncertainty in non-linear hydrological models using a Bayesian approach, in *Statistics for the Environment*, vol. 2, *Water Related Issues*, edited by V. Barnett and K. F. Turkman, pp. 297–317, John Wiley, Hoboken, N. J., 1994.

Romanowicz, R., K. J. Beven, and J. A. Tawn, Bayesian calibration of flood inundation models, in *Flood Plain Processes*, edited by M. G. Anderson and D. E. Walling, John Wiley, Hoboken, N. J., 1996.

Rothman, D. H., Automatic estimation of large residual statics, *Geophysics*, *51*, 332–346, 1986.

Sen, M. K., and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion*, 281 pp., Elsevier Sci., New York, 1995.

Sen, M. K., and P. L. Stoffa, Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion, *Geophys. Prospect.*, *44*, 313–350, 1996.

Sen, O. L., L. A. Bastidas, W. J. Shuttleworth, Z. L. Yang, H. V. Gupta, and S. Sorooshian, Impact of field-calibrated vegetation parameters on GCM climate simulations, *Q. J. R. Meteorol. Soc.*, *127*, 1199–1223, 2001.

Slater, A. G., et al., The representation of snow in land surface schemes: Results from PILPS 2(d), *J. Hydrometeorol.*, *2*, 7–25, 2001.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Sci., New York, 1987.

Thiemann, M., M. Trosser, H. Gupta, and S. Sorooshian, Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*, 2521–2535, 2001.

Wang, Q. J., The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, *Water Resour. Res.*, *27*, 2467–2471, 1991.

Xia, Y., A. J. Pitman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas, Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw data set, *J. Hydrometeorol.*, *3*, 181–194, 2002.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*, 83–97, 1998.

C. Jackson, M. K. Sen, P. L. Stoffa, and Y. Xia, Institute for Geophysics, The John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, 4412 Spicewood Spring Road, Building 600, Austin, TX 78759-8500, USA. (charles@ig.utexas.edu; mrinal@ig.utexas.edu; pauls@ig.utexas.edu; youlong@ig.utexas.edu)